Operons Form Because of Co-regulation and Not by Horizontal Gene Transfer

M.N. Price[1,4], K.H. Huang[1,4], A.P. Arkin[1,2,3,4], E.J. Alm[1, 4]

[1]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, [2]Howard Hughes Medical Institute, Berkeley, CA, USA, [3]Department of Bioengineering, UC Berkeley, USA
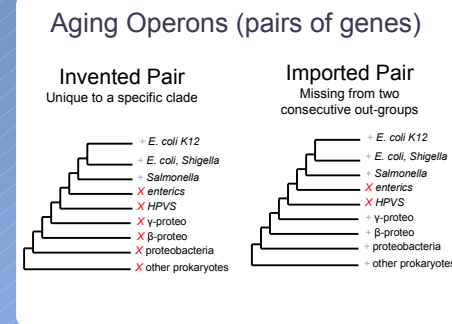[4]Virtual Institute for Microbial Stress and Survival

Abstract

Why Are There Operons?
- Co-regulation selects for new operons
  - Operons have more upstream regulatory information
- Operons are not selfish
  - Operons form without horizontal gene transfer (HGT)
    - Non-HGT genes are mostly in operons
    - Non-HGT and native genes form new operons
  - New operons are not functionally coherent
  - HGT does help maintain operons
    - Associated with transfer of existing operons, not with the creation of new operons
    - Insertion of new genes often creates operons

Theories of Operon Evolution

- Co-regulation
  - Clustering genes under the same promoter provides a selective advantage to the organism
- Selfish operons
  - Clustering of genes into operons makes them more likely to propagate by HGT

- Others: Thermoadaptation, Coevolution

Operons Have More Complex Regulation

Phylogenetic footprints in *E. coli*
(from McCue et al. 2002)
Only the first gene in predicted operons
Only genes with footprints

Aging Operons (pairs of genes)

Invented Pair — Unique to a specific clade

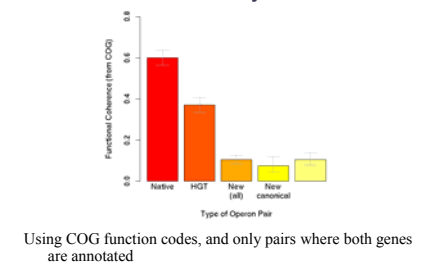Imported Pair — Missing from two consecutive out-groups
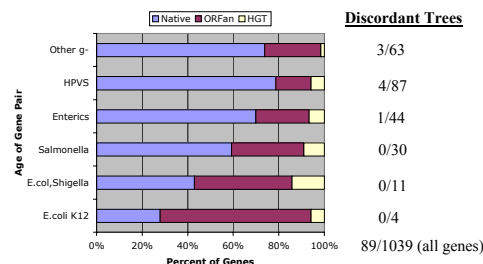

Aging Genes

Similar to Aging Operons
- Native to the proteobacteria
  - Homologs in every clade
- Non-HGT (special subset of Native)
  - Single-copy ubiquitous COGs
  - Gene tree agrees with species tree (Lerat et al.2003)
- ORFan
  - No homologs outside clade

New Operons Are Not as Functionally Coherent

Using COG function codes, and only pairs where both genes are annotated

Inventing Operons without HGT

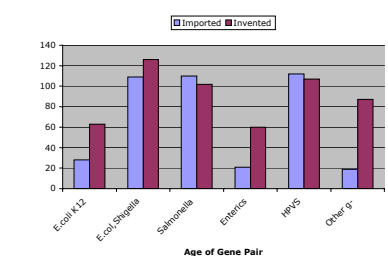| Age of Gene Pair	Discordant Trees
Other g-	3/63
HPVS	4/87
Enterics	1/44
Salmonella	0/30
E.coli,Shigella	0/11
E.coli K12	0/4
	89/1039 (all genes)

New Operons of Non-HGT Genes

- 37 new operons from 200 **non-HGT** genes
  - Known operons, or predictions confirmed by microarray data
- Many essential genes (*)
- Sometimes merge two pre-existing operons or add to one

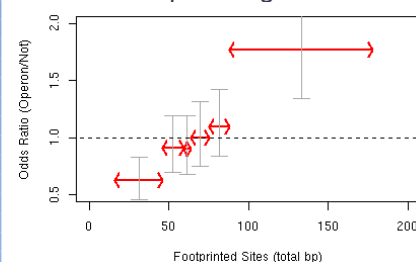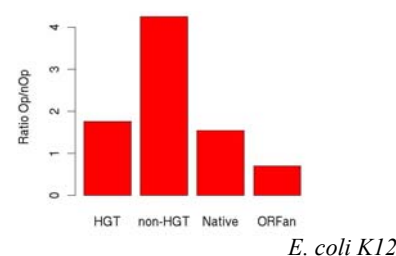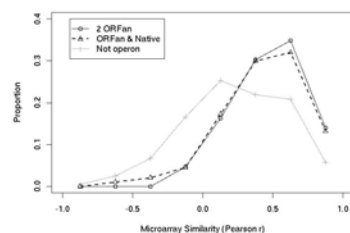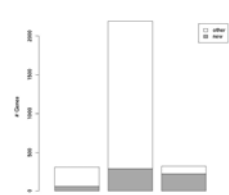| Age of Operon	Known Operon	Microarray similarity	Upstream Gene	Downstream Gene	Age of non-hold gene
Salmonella	No	0.51	yhB	yhfC *	Native
HPVS	No	0.76	ysrD	yrdC	Native
HPVS	No	0.80	yrdC	aroE	Native
HPVS	No	0.73	yhbE	ybbZ *	Native
HPVS	No	0.71	pyrF *	ycdI	Native
HPVS	No	0.67	murB	birA *	Native
HPVS	No	0.85	ygiM	cca *	Native
HPVS	Yes	0.61	kdtA *	kdtB *	Native
HPVS	No	0.58	ygg J	gshB *	Native
HPVS	No	0.54	yhbF *	yhhL *	gammas
HPVS	No	-0.13	pyrE	tpb	Native
gammas	Yes	0.81	lgt *	ftyA *	Native
gammas	Yes	0.76	lspA *	slpA	Native
gammas	No	0.92	rpsF	prlB *	beta/gamma
gammas	Yes	0.93	priB *	rpsR *	beta/gamma
gammas	No	0.87	nhpB	dapA *	beta/gamma
gammas	No	0.59	leqA	apaG	Native
gammas	No	0.72	fbJ	hfB	Native
gammas	Yes	0.69	yacE *	yacF	beta/gamma
gammas	No	0.64	folC	dcdD	Native
gammas	No	0.34	ycfC	purfB	gammas
beta/gamma	Yes	0.63	sdhB	sucA *	Native
beta/gamma	Yes	N/A	slpA	hflB *	Native
beta/gamma	Yes	0.92	rpsF	prlB *	beta/gamma
beta/gamma	Yes	0.93	priB *	rpsR *	beta/gamma
beta/gamma	No	0.87	nhpB	dapA *	beta/gamma
beta/gamma	No	0.85	atpI	atpB *	gammas
beta/gamma	Yes	0.72	fbJ	hfB	Native
beta/gamma	Yes	0.69	yacE *	yacF	beta/gamma
beta/gamma	No	0.64	folC	dcdD	Native
beta/gamma	No	0.54	ycfB *	dapE *	Native
proteo	Yes	0.98	sucB *	sucC	Native
proteo	Yes	0.85	rnc *	era *	Native
proteo	No	0.83	yaeL	yaeT *	Native
proteo	No	0.76	ydgQ	nth	Native
proteo	No	0.74	ndk	yfgB	Native
proteo	Yes	0.73	pdxA	lscgA	Native
proteo	No	0.64	pheT	himA	Native
proteo	Yes	0.63	nhF *	ileS *	N/A
proteo	No	0.52	b2512	b2511 *	Native

Many Operons Are Imported

- Imported operons are 2x enriched in HGT genes
- Gene's age usually agrees with operon's age (simple transfer)

Non-HGT Genes Are More Likely to Be in Operons

E. coli K12

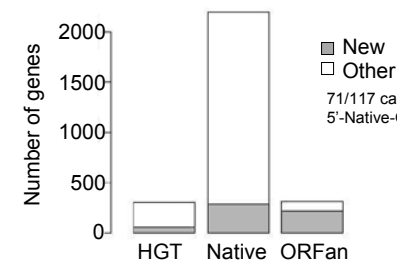## Operons Sometimes Form When ORFans Insert

Newest operons* often contain ORFans

Validation of new operons containing ORFans

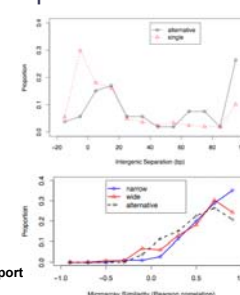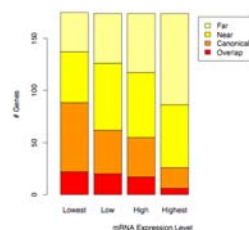* Unique to Enterobacteria

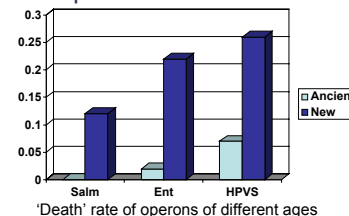## New Operons Contain Many ORFans

71/117 cases
5'-Native-ORFans-3'

Alternative Transcripts and Widely Spaced Operons Are Similar

47/156 'wide' operons have conserved footprints
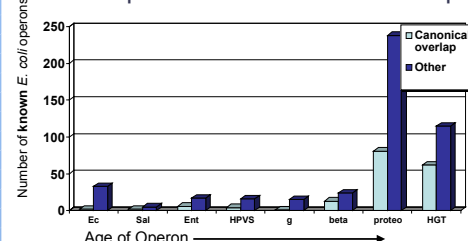**Most have support from literature**

Life Cycle of an Operon

- Birth
  - Gene pairs are merged into new operons
    - By rearrangement or deletion of intervening genes
  - Ancient (and non-HGT) genes can form operons
  - Useful operons are selected for advantageous regulation
  - Some new operons allow ORFans to be expressed?
- Death
  - Gene order is shuffled extensively across species
  - Operons involving highly conserved or essential genes resist disruption
    - Genes in operons tend to be more conserved & more tightly regulated
  - Immigration
    - Many operons are imported from distant lineages
    - Operons are a useful unit of HGT

New Operons are Less Stable

'Death' rate of operons of different ages

New Operons Have Fewer Overlaps

Highly Expressed Genes Have Longer Spacings

Some of these results just appeared in
M. N. Price et al., Genome Research 15:809-819